

Data Mining and Machine Learning in the Context of Sustainable Evaluation: A Literature Review

J. Souza, A. Francisco, C. Piekarski, G. Prado, and L. Oliveira

Abstract—Measuring and evaluating the sustainable performance of an organization has become an important and challenging topic because it involves the economic, social and environmental dimensions, helping the development of policies and becoming strategic factors in the decision-making process. However, difficulties are still encountered by managers in adequately assessing sustainability at the corporate level. In this perspective, data mining and machine learning are presented as techniques for extracting potentially useful information for generation of knowledge. Therefore, the purpose of this article is to identify, by means of a literature review, different approaches used to assist in the evaluation of sustainable performance. The method called *Methodi Ordinatio* was used for the review and, for the analysis, the software tools: VOSviewer e RStudio. By means of the methodological procedure adopted, 33 significant articles were identified for analysis from the Web of Science, Scopus and Science Direct databases, in which mainly the applied techniques were addressed. In this sense, this study seeks to stimulate research on the use of DM and ML to help in the sustainable context, being essential to assist in the Sustainable Development Goals.

Keywords— Data mining, Machine learning, Literature review, *Methodi Ordinatio*, Sustainable performance.

I. INTRODUÇÃO

PREOCUPAÇÕES sobre a sustentabilidade são crescentes e surgem devido à utilização inadequada de recursos naturais e à degradação progressiva dos ecossistemas, o que despertou o interesse de vários campos da ciência e de sistemas produtivos [1]–[3]. Isso ocorre durante o processo de desenvolvimento econômico de governos e empresas que, frequentemente, exploram recursos dos ecossistemas de maneira totalmente insustentável [4].

J. T. Souza é do Departamento de Pós-Graduação em Engenharia de Produção, da Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil (e-mail: jovanisouza5@gmail.com)

A. C. Francisco é do Departamento de Pós-Graduação em Engenharia de Produção, da Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil (e-mail: acfrancisco@utfpr.edu.br)

C. M. Piekarski é do Departamento de Pós-Graduação em Engenharia de Produção, da Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil (e-mail: piekarski@utfpr.edu.br)

G. F. Prado é do Departamento de Pós-Graduação em Engenharia de Produção da Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil (e-mail: guilhermefprado92@gmail.com)

L. G. Oliveira é do Departamento de Pós-Graduação em Engenharia de Produção, da Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil (e-mail: leandro.gasparello@gmail.com)

No entanto, as organizações estão cada vez mais motivadas a adaptar-se ao chamado *triple bottom line* (tripé da sustentabilidade), que considera os aspectos éticos e sociais dos negócios, bem como o advento da globalização, poluição ambiental e escassez de recursos [5], [6].

Neste contexto, a sustentabilidade tornou-se um conceito de grande importância para a governança corporativa das empresas, pois beneficia no desenvolvimento de políticas e no processo de tomada de decisões, aliada às questões sustentáveis [7]–[11].

Porém, ainda são necessárias investigações mais aprofundadas com relação a essa temática, especialmente no que se refere à avaliação de desempenho sustentável [10]. Analisar o desempenho sustentável de uma determinada empresa ainda é um desafio para a academia e a indústria [12].

O processo de desempenho sustentável está relacionado diretamente aos impactos agregados dos fatores econômicos, sociais e ambientais, sendo eles positivos ou negativos. Os autores salientam, também, que não se pode enfatizar em apenas uma dimensão [13].

As empresas e seus *stakeholders* estão interessados no desempenho ambiental e social de produtos e serviços, e não apenas em seu desempenho econômico [14]. Ademais, na prática, a sustentabilidade é geralmente reduzida ou substituída apenas por aspectos econômicos ou ambientais [15]. Estudos anteriores, porém, mostram que a busca constante por objetivos econômicos não pode levar à sustentabilidade [16], ou seja, os estudos voltados à sustentabilidade geralmente não consideram todo o seu tripé.

Com relação à medição da sustentabilidade, alguns critérios são necessários, como: ter um objetivo definido e conhecer a situação da empresa, a coleta dos dados, obter informações confiáveis e comparáveis, além da indicação dos problemas e possíveis soluções [17]–[19].

Além disso, o conjunto de indicadores que combinam várias dimensões da sustentabilidade ajudam as empresas a medir seus esforços de sustentabilidade em uma escala muito maior, comparando-os com o uso de indicadores individuais [20].

Nesta perspectiva, são necessárias metodologias e ferramentas capazes de modelar os diferentes impactos nas dimensões da sustentabilidade, apresentando corretamente o grau de correlação, permitindo assim, o envolvimento mais ativo das partes interessadas [21]–[23].

Contudo, nem todas as metodologias utilizadas para a sustentabilidade têm sido eficazes [24], [25], uma vez que não se preocupam igualmente entre as dimensões.

Outro fator preponderante para esse contexto é a necessidade de se utilizar métodos computacionais que auxiliem na avaliação adequada do desempenho sustentável das empresas, uma vez que a rapidez na veiculação das inúmeras informações, tem se tornado uma constante, devido ao desenvolvimento científico e tecnológico [26], [27]. A extração de informações em grandes conjuntos de dados requer diferentes técnicas analíticas [28].

Assim sendo, a aplicação de mineração de dados (*data mining* – DM) juntamente com aprendizado de máquina (*machine learning* – ML) são primordiais para a geração de conhecimento [29]–[32] em grandes volumes de dados, uma vez que por meio desses algoritmos, padrões, regras ou dados ocultos [33]–[35] são descobertos para serem aplicados em problemas reais [36], [37], auxiliando no processo de tomada de decisão [38], [39]. Em geral, o processo de mineração de dados combina algoritmos de ML e análises estatísticas na busca de similaridades ou até mesmo discordância entre os dados [40]. Nesse sentido, os algoritmos são aptos a analisar indicadores de desempenho, o que facilitaria o desenvolvimento de estudos relacionados à sustentabilidade [41].

Diante da importância e de acordo com as pesquisas realizadas, não há estudos publicados que relatem as diferentes abordagens envolvendo DM e ML para esta temática. Portanto, a contribuição deste trabalho é fornecer, a partir de uma revisão de literatura em revistas pertencentes à *Web of Science* (WoS), *Scopus* e *Science Direct*, as diferentes abordagens presentes em DM e ML para auxiliar na avaliação de desempenho sustentável. As revisões de literatura são primordiais, visto que desempenham papel crucial na síntese de resultados de pesquisas anteriores, fornecendo novos *insights* sobre a temática e avançando em uma determinada linha de pesquisa [42]–[46]. A metodologia denominada de *Methodi Ordinatio* [47] foi usada para encontrar artigos relevantes sobre o tema. Estabeleceu-se como limite temporal o período de 2000 a 2017.

O artigo está organizado da seguinte forma: a seção II apresenta os materiais e métodos utilizados na pesquisa. Os resultados e discussões são descritas na seção III e a seção IV apresenta a conclusão do artigo.

II. MATERIAIS E MÉTODOS

Com o objetivo de investigar informações essenciais relacionadas ao tema, uma revisão de literatura foi realizada. As etapas para esse processo foram feitas de acordo com a metodologia denominada de *Methodi Ordinatio* [47], que possui como base o modelo *ProKnow-C* [48]. A metodologia é composta por nove etapas. Posteriormente, foi utilizado o software *VOSviewer* [49] para ilustrar graficamente mapas de redes quanto à coocorrência das palavras-chaves, e o pacote *bibliometrix* [50] por meio do *RStudio* (ambiente integrado do software R), tendo como finalidade apresentar os artigos e países com mais citações, bem como a rede bibliográfica considerando os links de colaboração entre os países no estudo realizado.

A. Descrição das etapas da metodologia *Methodi Ordinatio*

Etapa 1: Estabelecendo a intenção da pesquisa

Refere-se ao objetivo em que o tema se baseia. Nesse caso, portanto, é identificar, a partir de uma revisão de literatura, as técnicas mais utilizadas para auxiliar no processo de avaliação de desempenho sustentável. Sendo assim, as questões de pesquisa foram divididas em: questão primária (QP) e questões secundárias (QS).

A seguir, são relatadas as questões:

QP: Como a utilização da mineração de dados e aprendizado de máquina estão sendo aplicadas para auxiliar na avaliação de desempenho sustentável?

QS1: Quais as técnicas/algoritmos mais comumente utilizadas nesta temática?

QS2: Qual é o panorama atual das pesquisas?

Etapa 2: Pesquisa preliminar nas bases de dados

Nessa etapa, são escolhidas as principais palavras-chaves para o estudo. Nota-se que a seleção de palavras-chave e a identificação de sua relação são fatores cruciais para fornecer resultados fidedignos, no entanto os conhecimentos adquiridos são escolhidos com base no julgamento dos autores [51], [52].

Para a etapa 2 é essencial utilizar um gerenciador de referências. Nesse estudo, empregou-se o *Mendeley*. O software *Mendeley* auxiliou na organização dos artigos encontrados, bem como na exclusão dos artigos duplicados. Para este estudo, foram utilizadas as seguintes palavras chaves: “*data mining*”, “*machine learning*”, “*data analytics*”, “*sustainable performance*”, “*sustainable evaluation*”, “*sustainable consumption*”, “*sustainable assessment*”, “*sustainability indicators*” e “*sustainability indices*”.

Etapa 3: Definição da combinação das palavras-chave e das bases de dados a serem utilizadas

Para esta etapa, as bases de dados e a combinação de palavras-chaves foram respectivamente: *Web of Science* (WoS), pela sua abrangência, multidisciplinaridade e indexando apenas os periódicos mais citados, considerando suas respectivas áreas. Segundo a WoS [53], existem mais de 12.000 periódicos de alto impacto, indexados nessa base. Ainda, foram utilizadas as bases *Scopus* e *Science Direct*. O limite temporal para o estudo foi de 2000 a 2017. Em relação às combinações, foram formados dois grupos:

- Grupo 1: Palavras relacionadas a DM e ML: “*data mining*”; “*machine learning*”; “*data analytics*”.
- Grupo 2: Palavras relacionadas a desempenho sustentável: “*sustainable performance*”; “*sustainable evaluation*”; “*sustainable consumption*”; “*sustainable assessment*”; “*sustainability indicators*”; “*sustainability indices*”.

Etapa 4: Busca final nas bases de dados

A quantidade de artigos encontrados através da Etapa 3, foram de 89 documentos, sendo 67 artigos de periódicos, 10 resumos de conferência, 7 artigos de conferência, 3 enciclopédias e 2 editoriais. Para esses valores, não foram contabilizados os critérios de exclusão dos artigos duplicados e os sem vínculos com o assunto.

Etapa 5: Procedimentos de filtragem

Os procedimentos de filtragem foram utilizados para eliminar os artigos repetidos, além daqueles artigos que não possuíam vínculo com a temática. Neste aspecto, o software *Mendeley* auxiliou no processo de exclusão de duplicatas. Em relação aos artigos que não estavam relacionados, foram lidos os títulos e resumos e, posteriormente, partiu-se para a leitura integral com o objetivo de confirmar o alinhamento com o assunto. Para este estudo, não foram considerados artigos de conferência, enciclopédia, livros e capítulos de livros, pois se buscou apenas periódicos com fator de impacto. O fator de impacto é uma medida usada para avaliar a relevância de um periódico ao longo do tempo [54] e é calculado a partir da média de citações por artigo para um periódico, no período de dois anos [55].

Após a filtragem, 33 artigos foram considerados para análise. É importante mencionar que foi utilizado o software *JabRef*, com o intuito de exportar os dados dos gerenciadores de referência para formato de planilha eletrônica [56].

Etapa 6: Identificação do fator de impacto, ano de publicação e número de citações

Para cada artigo, foi analisado o fator de impacto da revista (*Journal Citation Reports – JCR*), o ano de publicação, além do número de citações. Ambos os artigos foram organizados no ambiente do *Excel*.

Etapa 7: Classificando os artigos utilizando o *InOrdinatio*

Para definir a relevância dos artigos, foi utilizada a equação *InOrdinatio* [47]. A equação combina fator de impacto, ano de publicação e número de citações [56]. Ambos os artigos foram classificados por ordem de relevância, considerando as métricas mencionadas. O resultado final é apresentado através do Quadro I, de acordo com o nível de importância.

Etapa 8: Localizando os trabalhos em formato integral

Nesta etapa, foram localizados todos os artigos determinados já pela Etapa 5 (33 artigos), ou pelo site da revista ou pelo site do *Google Scholar*. É importante frisar, nesta etapa, que existem alguns títulos muito semelhantes aos dos artigos definidos ou escolhidos em revistas diferentes do identificado anteriormente, logo, é imprescindível atentar-se a essas questões, para evitar análises incoerentes e inadequadas.

Etapa 9: Leitura sistemática e análise de artigos

Por fim, na última etapa é realizada a leitura sistemática dos artigos, bem como sua análise. Destaca-se que a escolha de todos os artigos fica a critério do pesquisador, sendo que o mesmo poderá escolher os cinco ou dez mais relevantes, e assim, sucessivamente. Para este estudo, optou-se pela leitura e análise dos 33 artigos (Quadro I).

procedimento metodológico adotado (*Methodi Ordinatio*). Foram analisadas em cada artigo, principalmente, abordagens que contemplavam DM e ML, tais como: processos, métodos, técnicas, modelos, algoritmos, dentre outros. Ademais, informações de pesquisa foram fornecidas por meio de mapas visuais em relação a rede de coocorrência de palavras-chaves e também, foram apresentados os artigos e países com mais citações, bem como a rede bibliográfica considerando os links de colaboração entre os estudos investigados.

O Quadro I mostra os artigos em ordem de relevância, após os devidos procedimentos demonstrados na seção anterior. A primeira coluna relata o título do artigo. A segunda coluna demonstra o título do periódico no qual o artigo foi publicado, seguido pela terceira coluna que demonstra o ano de sua publicação. A última coluna demonstra o fator de impacto (JCR) de cada um dos periódicos envolvidos.

TABELA I
RESULTADO FINAL APÓS A UTILIZAÇÃO DO *METHODI ORDINATIO*

Título do artigo	Título do periódico	Ano	JCR
The impact of big data on world-class sustainable manufacturing [57]	The International Journal of Advanced Manufacturing Technology	2016	2,601
Supplier selection for sustainable operations: A triple-bottom-line approach using a Bayesian framework [58]	International Journal of Production Economics	2015	4,407
Sustainability assessment of US manufacturing sectors: an economic input output-based frontier approach [59]	Journal of Cleaner Production	2013	5,651
Environmental, social and economic information management for the evaluation of sustainability in urban areas: A system of indicators for Thessaloniki, Greece [60]	Cities	2010	2,704
Evaluating energy performance in non-domestic buildings: A review [61]	Energy and Buildings	2016	4,457
Food safety pre-warning system based on data mining for a sustainable food supply chain [62]	Food Control	2017	3,667
Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry [63]	International Journal of Production Economics	2014	4,407
Evaluation of energy efficiency measures sustainability by decision tree method [64]	Energy and Buildings	2014	4,457
Urban sustainability management: A deep learning perspective [65]	Sustainable Cities and Society	2017	3,073
Text mining-based categorization and user perspective analysis of environmental sustainability indicators for manufacturing and service systems [66]	Ecological Indicators	2017	3,983
Sustainability trends in the process industries: A text mining-based analysis [67]	Computers in Industry	2014	2,850
Identification of key energy efficiency drivers through global city benchmarking: A data driven approach [68]	Applied Energy	2017	7,900
Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach [69]	Computer Methods and Programs in Biomedicine	2017	2,674
Integrating agricultural sustainability into policy planning: A geo-referenced framework based on Rough Set theory [70]	Environmental Science & Policy	2015	3,826
Identifying services for short-term load forecasting using data driven models in a Smart City platform [71]	Sustainable Cities and Society	2017	3,073

III. RESULTADOS E DISCUSSÃO

Com o intuito de responder as questões de pesquisas feitas previamente, uma revisão de literatura foi empregada. Foram tomadas medidas durante a seleção desses artigos para permanecerem consistentes com o problema de pesquisa. Neste sentido, foram selecionados 33 artigos seguindo o

Decision-making framework with double-loop learning through interpretable black-box machine learning models [72]	Industrial Management & Data Systems	2017	2,948
A differential multi-criteria analysis for the assessment of sustainability performance of European countries: Beyond country ranking [73]	Journal of Cleaner Production	2017	5,651
The dynamics of public safety in cities: A case study of Shanghai from 2010 to 2025 [74]	Habitat International	2017	3,000
Understanding the effects of environmental factors on building energy efficiency designs and credits Case studies using data mining and real-time data [75]	Journal of Engineering Design and Technology	2017	0,5900
Intelligent data analysis for sustainable smart grids using hybrid classification by genetic algorithm based discretization [76]	Intelligent Decision Technologies- Netherlands	2017	0,333
Accurate prediction of sugarcane yield using a random forest algorithm [77]	Agronomy for Sustainable Development	2016	4,503
Analyzing sustainability of Chinese mining cities using an association rule mining approach [78]	Resources Policy	2016	2,695
Data mining for energy analysis of a large data set of flats [79]	Proceedings of the Institution of Civil Engineers - Engineering Sustainability	2016	1,500
Sustainable service and energy provision based on agile rule induction [80]	International Journal of Production Economics	2016	4,407
A text mining framework for advancing sustainability indicators [81]	Environmental modelling & software	2014	4,177
Modelling forest growing stock from inventory data: A data mining approach [82]	Ecological Indicators	2014	3,983
Sustainable Road Management in Texas: Network-Level Flexible Pavement Structural Condition Analysis Using Data-Mining Techniques [83]	Journal of Computing in Civil Engineering	2012	2,310
Monitoring Global Forest Cover Using Data Mining [84]	ACM Transactions on Intelligent Systems and Technology	2011	3,190
Estimating the risk of fire outbreaks in the natural environment [85]	Data Mining and Knowledge Discovery	2012	2,481
Mining sustainability indicators to classify hydrocarbon development [86]	Knowledge-Based Systems	2011	4,396
Evaluation of sustainability using fuzzy association rules mining [87]	The International Journal of Advanced Manufacturing Technology	2011	2,601
Searching for simplified farmers' crop choice models for integrated watershed management in Thailand: A data mining approach [88]	Environmental Modelling & Software	2009	4,177
Use of Rough Sets analysis to classify Siberian forest ecosystems according to net primary production of phytomass [89]	Infor: Information Systems and Operational Research	2000	0,257

Os trabalhos abaixo serão analisados de acordo com a ordem apresentada no Quadro I. O primeiro artigo investigado [57] analisa a importância do *big data* no suporte à manufatura sustentável. Análise de componentes principais (*Principal component analysis* – PCA) e projeção aleatória (*Random project*) foram utilizadas no conjunto de dados coletados, com o objetivo de identificar fatores importantes para construção de um *framework* conceitual.

Uma abordagem baseada na análise *bayesiana* juntamente com a simulação de *Monte Carlo* para a avaliação e classificação de fornecedores sustentáveis para cadeia de suprimentos também é introduzida [58], considerando múltiplas dimensões do tripé da sustentabilidade.

Uma ferramenta analítica que pode ser usada para analisar e comparar o desempenho de sustentabilidade dos setores de manufatura dos EUA também é apresentada [59]. Neste sentido, uma análise de mineração de dados foi realizada com o propósito de determinar os principais direcionadores dos setores da cadeia de suprimentos em termos de categorias de impacto.

Um sistema de indicadores foi desenvolvido e utilizado como uma ferramenta dinâmica para o gerenciamento de informações ambientais, sociais e econômicas, a fim de avaliar a sustentabilidade em áreas urbanas [60]. Neste contexto, a mineração de dados auxiliou na previsão de dados faltantes para o cálculo dos indicadores.

Autores também buscaram fazer revisões de literatura [61], na qual foram analisadas metodologias para análise de desempenho energético em edifícios não residenciais. Uma das metodologias empregadas se refere ao aprendizado de máquina (*machine learning*), mais precisamente, as redes neurais artificiais e análises de *cluster*. Em alguns estudos, SVM e árvore de decisão também foram implementados para avaliação do desempenho energético dos edifícios, porém não sendo tão difundidos quanto os anteriormente mencionados.

Um sistema de pré-aviso de segurança alimentar foi criado [62], adotando regras de associação e a tecnologia denominada de Internet das Coisas (*Internet of Things* – IoT), para monitorar todos os dados de detecção de toda a cadeia de suprimentos. O objetivo buscou melhorar a sustentabilidade da qualidade da cadeia de fornecimento.

Além disso, foi apresentado um sistema de apoio à decisão de sustentabilidade da qualidade da cadeia de suprimentos, adotando a mineração de regras de associação e a regra de combinação de técnicas de *Dempster* [63]. Ambas as técnicas são empregadas para explorar os bons planos de logística usados para transportar produtos alimentícios na rede de distribuição, de modo a reduzir a incerteza e gerenciar riscos na cadeia de fornecimento.

Outros autores [64] apresentam um método de suporte baseado em árvore de decisão, com o objetivo de ter uma estrutura clara e objetiva para a tomada de decisão contínua e de avaliação das medidas de eficiência energética. Cinco critérios que cobrem os aspectos de sustentabilidade foram considerados: eficiência energética, impacto ambiental, racionalidade econômica, conforto físico e duração sob o ponto de vista do ciclo de vida.

Análises de conceito formal (*Formal Concept Analysis*) foram empregadas [65], as quais têm como princípio regras de associação, para identificar pontos cruciais no desenvolvimento da sustentabilidade urbana, bem como aprendizagem profunda (*deep learning*) para avaliação de dados textuais obtidos *Carbon Disclosure Project* (CDP), organização que divulga informações a respeito de quatro áreas principais: riscos e oportunidades, estratégias e desempenho, emissões de gases de efeito estufa e governança.

A falta de orientação técnica e teórica e a utilidade dos indicadores de sustentabilidade ambiental para sistemas de manufatura e de serviços foram analisadas [66]. Para isso, 55 indicadores de sustentabilidade foram extraídos da literatura

existente. Técnicas de mineração de texto (modelos de tópicos) foram utilizadas, baseadas em algoritmos de aprendizado de máquina, a fim de categorizar os indicadores, a partir da extração de informações e similaridades conceituais entre os mesmos.

Ademais, as tendências e práticas de sustentabilidade nas indústrias de processos químicos foram investigadas [67], por meio de técnicas de mineração de texto em relatórios de sustentabilidade. O objetivo da técnica é extrair informações e tendências a partir da grande quantidade de dados textuais descritos nos relatórios.

Outros autores [68] identificaram, a partir de um estudo sistemático, os principais impulsionadores de eficiência energética urbana. Foram coletados dados de 25 cidades localizadas em três continentes: Ásia, Europa e América, por meio do relatório *Global Power City Index* (GPCI) e, também, por outras fontes. Para análise dos dados, foram utilizadas técnicas de *clustering*, árvore de decisão e análise envoltória de dados. A finalidade dessas técnicas é o cálculo e comparação da eficiência energética das cidades especificadas.

Investigou-se, também, a utilização de árvore de decisão (algoritmo J48) e de indução de regra (algoritmo PART), para dados demográficos e de pesquisa de saúde, na Etiópia [69]. O objetivo dessas técnicas foram de desenvolver um modelo de previsão para identificar áreas relacionadas à mortalidade infantil (indicador social).

Uma estrutura georreferenciada para avaliação de sustentabilidade agrícola foi proposta [70], visando apoiar o planejamento de políticas. Neste sentido, *Rough Set* é usada para redução de dados e indução de informações.

O modelo autorregressivo (AR) foi utilizado para prever carga de curto prazo em *Smart City*, caracterizando serviços que auxiliam na eficiência energética [71].

Outros autores [72] abordaram problemas com a aceitação de modelos de aprendizado de máquina em diferentes tipos de negócios. Para isso, os autores desenvolveram um *framework* adaptado aos modelos de aprendizado, a fim de formalizar processos de utilizar dados históricos da empresa, criando uma base para o aprendizado organizacional sustentável. Para a classificação foram utilizados os algoritmos: floresta aleatória; *naive bayes*; árvore de decisão; SVM (*support vector machines*). Já para a avaliação de desempenho e comparação dos modelos, utilizaram-se as métricas de acurácia da classificação e área sob a curva ROC.

Métodos de mineração de dados multivariados: PROMETHEE e DMA (*Differential multi-criteria analysis*) foram empregados [73], com o objetivo de analisar o desempenho sustentável referente à Estratégia de Desenvolvimento Sustentável da União Europeia, quanto às diferenças dos progressos entre os países membros, bem como avaliar o progresso no que se refere à redução do desenvolvimento desigual entre os países.

Investigou-se [74] a segurança pública urbana da cidade de Xangai, China, com a finalidade de construir um modelo dinâmico que imitasse o circuito real de segurança pública da cidade. Para isso, foram coletadas informações durante o período de 2000 a 2009, a respeito do desenvolvimento,

infraestrutura, habitabilidade, população, crime e desastre na cidade. As seguintes técnicas foram utilizadas: séries temporais, ajuste de curvas e regressão multivariada. O objetivo é garantir que o modelo construído se ajuste aos dados históricos.

Questões ambientais e o consumo de energia foram analisadas [75], mais especificamente a ligação entre temperatura atmosférica externa e energia consumida em edifícios reconhecidos pela LEED (Liderança em Energia e Design Ambiental), a partir de dados fornecidos em tempo real. Em vista disso, uma análise de *cluster* foi utilizada, com o objetivo de excluir fatores que poderiam influenciar os dados de consumo dos edifícios e, também, qui-quadrado para verificação dos resultados.

Técnicas de mineração de dados foram empregadas [76]: *best first tree* (BFTree), *genetic algorithm based discretization* (GAD), *ordered weighted average fuzzy-rough K-nearest neighbor* (OWAKNN) e *radial basis function* (RBF) *kernel* para previsão do comportamento dos clientes referente ao consumo energético voltado à sustentabilidade.

Florestas aleatórias de classificação e regressão a partir de dados climáticos e de produtividade coletados de 1992 a 2013 foram aplicados, para gerar um modelo de previsão, com o intuito de oferecer novos *insights* sobre a produtividade de cana-de-açúcar [77]. O estudo foi realizado no nordeste da Austrália.

Outro estudo [78] analisaram dados de 110 cidades de mineração localizadas na China, nas quais utilizaram regras de associação para identificar características e padrões, referentes às condições ambientais, sociais e econômicas das cidades, a fim de fornecer uma base para formulação de políticas.

Outros pesquisadores [79] conduziram a aplicação de uma metodologia baseada em mineração de dados para definir regras de tomada de decisão para identificar padrões de consumo de energia em um grande conjunto de dados relacionados a apartamentos (92.906). O algoritmo de árvore de classificação e regressão (CART) foi utilizado com o intuito de classificar as observações em um número finito de classes de consumo de energia. O algoritmo *K-means* auxiliou no agrupamento (*cluster*) das amostras identificadas como “alto consumo”.

Uma abordagem foi proposta [80] envolvendo *Rough Sets* para selecionar atributos significativos e induzir regras de decisão a partir de um conjunto de dados relevantes para serviços sustentáveis e fornecimento de energia.

A viabilidade de identificar, rastrear e relatar indicadores de sustentabilidade, a partir da análise de notícias digitais foi explorada em outro estudo [81]. A metodologia se deu a partir da combinação de técnicas de mineração de texto para classificar com maior precisão artigos de sustentabilidade. Árvores de categorias hierárquicas foram utilizadas para auxiliar no processo, além de GDA (*Generalized Discriminant Analysis*), para reduzir a dimensionalidade dos dados, KNN (*K-nearest neighbors*) e *dataless* para classificação dos documentos relacionados ou não relacionados aos indicadores de sustentabilidade.

Uma nova abordagem é apresentada para previsão de

estoque em crescimento (indicador ecológico) de florestas na Eslovênia, a partir de métodos de aprendizado de máquina e mineração de dados [82]. Diferentes abordagens de árvores de decisão foram aplicadas aos dados disponíveis pelos inventários florestais nacionais, a fim de determinar as principais variáveis que influenciam a dinâmica temporal do indicador, bem como prever o crescimento.

Outros autores [83] desenvolveram um método de predição baseado em mineração de dados, com intuito de determinar a condição estrutural de pavimento em níveis de rede. Para a escolha dos atributos-chave, foram utilizados algoritmos de classificação: *correlation-based feature selection* (CFS); *wrappers for feature subset selection algorithms*; *information-gain attribute ranking* (Infogain) e *symmetrical uncertainty attribute evaluators*.

Uma visão geral das oportunidades e desafios no desenvolvimento e na aplicação de abordagens de mineração de dados para fornecer informações críticas para o monitoramento do uso da floresta e da terra foi explorada [84]. Análise de séries temporais e associações foram aplicadas, a fim de detectar uma variedade de mudanças no ecossistema global.

Modelos preditivos foram construídos [85], a fim de estimar risco de incêndios na Eslovênia, a partir de dados do SIG (Sistema de Informação Geográfica), imagens de sensoriamento remoto e modelo de previsão do tempo, denominado ALADIN. Para tal, foram utilizados algoritmos de mineração de dados, tais como: *Bagging* e Floresta Aleatória. Para avaliação dos modelos, utilizaram-se as métricas de precisão e *recall*.

Outros autores propuseram um método de classificação para avaliar o desenvolvimento energético de um país a partir da utilização de técnicas de mineração de dados e indicadores de sustentabilidade [86]. O método baseou-se no *K-means clustering*, no qual não há um valor de classe. Portanto, o trabalho desenvolveu um método único para rotular classes não supervisionadas, as quais foram divididas em: futurista, em conformidade, sustentável, insustentável ou crítico. Posteriormente, foi utilizado um algoritmo de classificação de árvore de decisão denominado ID3 para simular os indicadores de sustentabilidade em uma estrutura hierárquica para facilitar a visualização ao avaliar o desenvolvimento.

A mineração de regras de associação *fuzzy* (*fuzzy association rules mining* – FARM) foi aplicada [87] para avaliar a sustentabilidade em uma organização indiana de fabricação de comutadores rotativos. O objetivo com essa técnica é auxiliar os tomadores de decisão de maneira mais precisa em situações complexas.

Utilizou-se o algoritmo C4.5, que baseia-se no modelo de árvore de decisão, para gerar regras de decisão simplificadas para a escolha de culturas dos agricultores em duas bacias hidrográficas localizadas na Tailândia [88]. Os fatores que influenciaram as decisões dos agricultores foram apresentados, bem como as variáveis utilizadas como classificadoras do modelo de escolha.

Atributos essenciais foram identificados para o desenvolvimento de práticas sustentáveis de manejo floresta

na Sibéria [89]. Em razão disso, os autores utilizaram métodos baseados na teoria RS (*Rough Sets*), que envolve estudar problemas de classificação. Foram geradas regras que possibilitaram identificar características comuns para escores pertencentes à mesma classe.

Enfatiza-se, a partir dos trabalhos mencionados, que as técnicas podem ser utilizadas em diferentes situações, como para previsão, filtragem, comparação, análise, identificações de padrões, classificações, dentre outras utilidades. O mais comumente utilizado nestes estudos foi a utilização de análises preditivas (QP). Árvores de decisão, regras de associação e análises de *cluster* auxiliaram na maioria dos estudos analisados (QS1).

As árvores de decisão são modelos supervisionados que podem ser usados tanto para classificação como para regressão [90], [91]. Porém, são necessários seguir algumas diretrizes para sua utilização, tais como: ordenar de forma correta os dados; descobrir qual estrutura mais se adequa ao seu propósito; entender as variáveis inseridas no contexto e construir um modelo auto-adaptável [91]. Além disso, têm preferência dentre as outras técnicas, pois permitem aos usuários entender facilmente o comportamento dos modelos gerados [92].

Já as regras de associação identificam relação entre itens frequentes em bancos de dados [93], [94], fornecendo suporte para identificação de conhecimento novo, potencialmente útil e na forma de implicação $X \rightarrow Y$ [95]. O processo de regras de associação é constituído basicamente de duas etapas: encontrar todos os conjuntos de itens frequentes que atendem ao limite mínimo de suporte (frequência que os padrões ocorrem) e gerar regras de associação fortes a partir de uma precisão mínima exigida de confiança [34].

Análises de *cluster* são utilizadas para agrupar objetos em um conjunto de dados com o intuito de que os componentes daquele grupo sejam semelhantes entre si, heterogêneo dos elementos agrupados em outros *clusters* [96], [97]. A validade dos resultados produzidos depende diretamente sobre que tipo de *cluster* irá utilizar e qual abordagem escolher para um conjunto final de *clusters* [98].

Em geral, de acordo com as pesquisas realizadas, esta é a primeira vez que um estudo foi conduzido para identificar diferentes abordagens de DM e ML para auxiliar na avaliação de desempenho sustentável. Nesse sentido, uma solução prática foi realizar uma revisão de literatura para encontrar essas respostas. Um estudo minucioso foi exigido pelos autores para isso. Observou-se, portanto, que análises preditivas foi o mais comumente utilizado. Predições são geralmente escolhidas pois podem ser utilizadas em uma gama de situações, além de apresentarem boa resposta preditiva e de fácil interpretação [99]. Árvore de decisão, regras de associação e análises de *cluster* auxiliaram na maioria dos estudos investigados.

Recomenda-se para estudos futuros, encorajar novas pesquisas sobre os reais impactos gerados a partir das abordagens estudadas (DM e ML) no contexto sustentável. Podem ser analisados, também, de forma mais precisa, as diferentes abordagens utilizadas pelos artigos estudados.

Por fim, é necessário que a escolha da técnica apropriada seja feita de forma criteriosa, já que o conhecimento que será adquirido depende diretamente da avaliação e análise das referidas técnicas. Além disso, podem ser altamente benéficas para auxiliarem no cumprimento dos Objetivos de Desenvolvimento Sustentável (ODS), através da geração de dados confiáveis e relevantes.

A. Análise de Coocorrência e Citações no Estudo Realizado

Para análise das palavras-chave foi utilizada a coocorrência, como apresentado na Fig.1. A coocorrência refere-se à relação de itens com base no número de documentos em que eles ocorrem juntos [49], abordando os pontos potenciais (*hotspots*) de pesquisas nos campos da disciplina, na qual fornece suporte auxiliar para a pesquisa científica [100].

O tamanho dos nós e das palavras representam os pesos dos nós, ou seja, quanto maior o nó e a palavra, maior o peso [101]. A distância entre os nós reflete a força de relação entre eles. Quanto mais espessa a linha, maior a coocorrência [100], [102]. Nota-se pela figura acima, portanto, que as palavras com maior frequência e força total do link dos nós (soma do link desse nó sobre todos os outros nós) [101] foram: *data mining*, *sustainability*, e, em seguida, *decision trees* (árvores de decisão). Outro fator preponderante é que os estudos designados concentram-se nos anos mais recentes, como consta na figura acima, indicando assim, que as análises para essa temática são relevantes, atuais e precisam de novos aprofundamentos (QS2).

Em relação às citações, a Tabela I e a Tabela II apresentam os dez artigos com o maior número de citações, além do total de citações por países, respectivamente.

TABELA I
TOP 10 – ARTIGOS COM MAIS CITAÇÕES

Artigo	Total de Citações (TC)	TC por ano
Egilmez G, 2013, J Clean Prod [57]	90	18.00
Dubey R, 2016, Int J Adv Manuf Techno [55]	65	32.50
Moussiopoulos N, 2010, Cities [58]	31	8.12
Sarkis J, 2015, Int J Prod Econ [56]	27	21.33
Ting Sl, 2014, Int J Prod Econ [61]	26	7.75
Mikucioniene R, 2014, Energy Build [62]	24	6.75
Borgstein Eh, 2016, Energy Build [59]	21	13.00
Liew Wt, 2014, Comput Ind [66]	16	5.25
Mithal V, 2011, Acm Trans Intell Syst Technolog [82]	15	2.71
Wang J, 2017, Food Control [60]	13	16.00

TABELA II
TOP 10 - TOTAL DE CITAÇÃO POR PAÍSES

Países	Total de Citações
Estados Unidos	136
Grécia	65
Reino Unido	31
Lituânia	27
Brasil	26
China	24
Singapura	21
Austrália	16
Paquistão	15
Eslovênia	13

Nesta perspectiva, a Fig. 2 ilustra uma rede bibliográfica considerando os links de colaboração entre países.

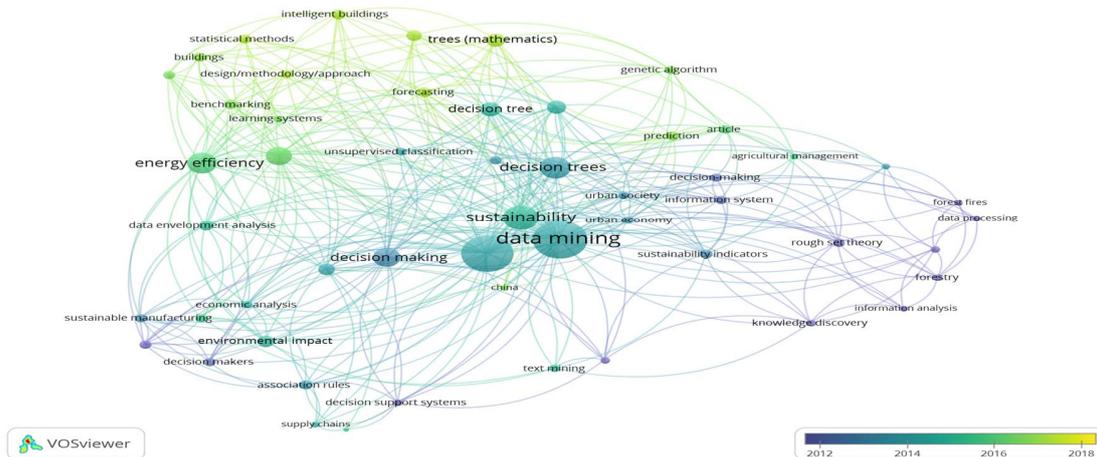


Fig. 1. Coocorrência de palavras-chave no estudo realizado.

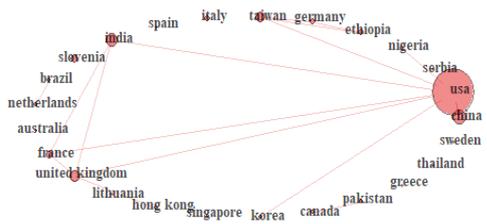


Fig. 2. Rede de colaboração entre os países no estudo realizado.

Os dados apresentados na Fig. 2 mostram uma forte conexão entre alguns países e os Estados Unidos, principalmente, a China. A Grécia, mesmo estando em segundo lugar no que se refere ao número total de citações (Tabela II), não possui vínculo com nenhum outro.

Por fim, as principais conclusões do trabalho serão apresentadas a seguir.

IV. CONCLUSÃO

O presente trabalho buscou contribuir, através de uma revisão da literatura, na identificação de diferentes abordagens utilizadas na mineração de dados e aprendizado de máquina para auxiliar na avaliação de desempenho sustentável. Para isso, se fizeram necessárias a utilização de uma metodologia de revisão e, também, a escolha das bases de dados que, neste estudo, definiu-se como: *Methodi Ordinatio*, *Web of Science*, *Scopus* e *Science Direct*, respectivamente. Estabeleceu-se como limite temporal o período de 2000 a 2017.

A partir do trabalho realizado, inúmeras utilidades foram percebidas com a utilização de DM e ML, tais como: análises preditivas (a grande maioria nesse estudo); filtro; comparação; análise; classificação; identificação de padrões, dentre outros. Destaca-se, também, que as técnicas mais empregadas foram com a utilização de árvores de decisão, regras de associação e análises de *cluster* para esses estudos. Além disso, os estudos encontrados concentram-se nos anos mais recentes, ou seja, são atuais e podem ser aprofundados para futuras pesquisas.

Indiferentemente do setor analisado, a seleção de uma técnica adequada é primordial para o entendimento das informações geradas, bem como para suporte ao processo decisório. Nesta perspectiva, este artigo pode contribuir para servir de base para estimular pesquisas sobre os reais impactos que os métodos abordados (DM e ML) podem fornecer para o âmbito sustentável. Além disso, mostrar que a discussão sobre essa temática é atual, importante e fundamental, uma vez que a sustentabilidade está cada vez mais participativa na governança corporativa de uma empresa. Além disso, é de grande interesse para os Objetivos de Desenvolvimento Sustentável (ODS), uma vez que priorizam a utilização de métodos computacionais para gerarem informações relevantes e de alta qualidade.

Para estudos futuros, podem ser analisados de forma mais específica, as diferentes técnicas utilizadas pelos artigos mencionados, bem como analisar outros artigos, em diferentes bases de dados.

REFERÊNCIAS

- [1] J. K. Hall, G. A. Daneke, and M. J. Lenox, "Sustainable development and entrepreneurship: Past contributions and future directions," *J. Bus. Ventur.*, vol. 25, no. 5, pp. 439–448, 2010.
- [2] C. Piekarski, A. Francisco, L. da Luz, T. Alvarenga, and J. Bittencourt, "Environmental profile analysis of MDF panels production: study in a Brazilian technological condition," *Cerne*, vol. 20, pp. 409–418, 2014.
- [3] E. Pinheiro, A. C. de Francisco, C. M. Piekarski, and J. T. de Souza, "How to identify opportunities for improvement in the use of reverse logistics in clothing industries? A case study in a Brazilian cluster," *J. Clean. Prod.*, vol. 210, pp. 612–619, 2019.
- [4] M. Song *et al.*, "How would big data support societal development and environmental sustainability? Insights and practices," *J. Clean. Prod.*, vol. 142, pp. 489–500, 2017.
- [5] A. Muller and A. Kolk, "Extrinsic and Intrinsic Drivers of Corporate Social Performance: Evidence from Foreign and Domestic Firms in Mexico," *J. Manag. Stud.*, vol. 47, pp. 1–26, 2010.
- [6] F. N. Ho, H.-M. D. Wang, and S. J. Vitell, "A Global Analysis of Corporate Social Performance: The Effects of Cultural and Geographic Environments," *J. Bus. Ethics*, vol. 107, no. 4, pp. 423–433, 2012.
- [7] M. Wagner, "The role of corporate sustainability performance for economic performance: A firm-level analysis of moderation effects," *Ecol. Econ.*, vol. 69, no. 7, pp. 1553–1560, 2010.
- [8] J. de Melo Junior, F. de Souza Farias, and O. Ryohei Kato, "Assessment of the Sustainability of Agroecosystems in the Amazon Region Using Neural Artificial Networks," *IEEE Lat. Am. Trans.*, vol. 14, no. 8, pp. 3804–3810, 2016.
- [9] S. Manzhynski, N. Siniak, A. Żróbek-Rózańska, and S. Żróbek, "Sustainability performance in the Baltic Sea Region," *Land use policy*, vol. 57, pp. 489–498, 2016.
- [10] P. Ahi, C. Searcy, and M. Y. Jaber, "A Quantitative Approach for Assessing Sustainability Performance of Corporations," *Ecol. Econ.*, vol. 152, pp. 336–346, 2018.
- [11] C. Xiao, Q. Wang, T. van der Vaart, and D. P. van Donk, "When Does Corporate Sustainability Performance Pay off? The Impact of Country-Level Sustainability Performance," *Ecol. Econ.*, vol. 146, pp. 325–333, 2018.
- [12] J.-P. Schögl, M. M. C. Fritz, and R. J. Baumgartner, "Toward supply chain-wide sustainability assessment: a conceptual framework and an aggregation method to assess supply chain performance," *J. Clean. Prod.*, vol. 131, pp. 822–835, 2016.
- [13] G. Büyüközkan and Y. Karabulut, "Sustainability performance evaluation: Literature review and future directions," *J. Environ. Manage.*, vol. 217, pp. 253–267, 2018.
- [14] M. Andersen and T. Skjoett-Larsen, "Corporate social responsibility in global supply chains," *Supply Chain Manag. An Int. J.*, vol. 14, no. 2, pp. 75–86, 2009.
- [15] M. Rodrigue, M. Magnan, and C. H. Cho, "Is Environmental Governance Substantive or Symbolic? An Empirical Investigation," *J. Bus. Ethics*, vol. 114, no. 1, pp. 107–129, 2013.
- [16] M. Song and Y. Zhou, "Quantitative Analysis of Foreign Trade and Environmental Efficiency in China," *Emerg. Mark. Financ. Trade*, vol. 52, no. 7, pp. 1647–1660, 2016.
- [17] A. A. da S. Batista and A. C. de Francisco, "Organizational Sustainability Practices: A Study of the Firms Listed by the Corporate Sustainability Index," *Sustainability*, vol. 10, no. 1, 2018.
- [18] A. Moldavska and T. Welo, "On the Applicability of Sustainability Assessment Tools in Manufacturing," *Procedia CIRP*, vol. 29, pp. 621–626, 2015.
- [19] A. Moldavska and T. Welo, "Development of Manufacturing Sustainability Assessment Using Systems Thinking," *Sustainability*, vol. 8, 2016.
- [20] C. B. Joung, J. Carrell, P. Sarkar, and S. C. Feng, "Categorization of indicators for sustainable manufacturing," *Ecol. Indic.*, vol. 24, pp. 148–157, 2013.
- [21] Q. Jiang *et al.*, "A principal component analysis based three-dimensional sustainability assessment model to evaluate corporate sustainable performance," *J. Clean. Prod.*, vol. 187, pp. 625–637, 2018.
- [22] E. D. R. S. Gonzalez *et al.*, "Making real progress toward more sustainable societies using decision support models and tools: introduction to the special volume," *J. Clean. Prod.*, vol. 105, pp. 1–13, 2015.
- [23] R. Caiado, G. Lima, L. Gavião, O. Quelhas, and F. Fidelis Paschoalino, "Sustainability Analysis in Electrical Energy Companies by Similarity Technique to Ideal Solution," *IEEE Lat. Am. Trans.*, vol. 15, no. 4, pp.

- 675–681, 2017.
- [24] R. Gahin, V. Veleva, and M. Hart, “Do Indicators Help Create Sustainable Communities?,” *Local Environ.*, vol. 8, no. 6, pp. 661–666, 2003.
- [25] S. Krank and H. Wallbaum, “Lessons from seven sustainability indicator programs in developing countries of Asia,” *Ecol. Indic.*, vol. 11, pp. 1385–1395, 2011.
- [26] R. Dubey *et al.*, “Can big data and predictive analytics improve social and environmental sustainability?,” *Technol. Forecast. Soc. Change*, 2017.
- [27] M.-L. Song, R. Fisher, J.-L. Wang, and L.-B. Cui, “Environmental performance evaluation with big data: theories and methods,” *Ann. Oper. Res.*, vol. 270, no. 1, pp. 459–472, Nov. 2018.
- [28] L. Duan and Y. Xiong, “Big data analytics and business analytics,” *J. Manag. Anal.*, vol. 2, no. 1, pp. 1–21, 2015.
- [29] R. Bonidia, J. Brancher, and R. Busto, “Data Mining in Sports: A Systematic Review,” *IEEE Lat. Am. Trans.*, vol. 16, no. 1, pp. 232–239, 2018.
- [30] S. M. M. Rubiano and J. A. D. García, “Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance,” *IEEE Lat. Am. Trans.*, vol. 14, no. 6, pp. 2783–2788, 2016.
- [31] D. Heredia, Y. Amaya, and E. Barrientos, “Student Dropout Predictive Model Using Data Mining Techniques,” *IEEE Lat. Am. Trans.*, vol. 13, no. 9, pp. 3127–3134, 2015.
- [32] J. T. de Souza, A. C. de Francisco, C. M. Piekarski, and G. F. do Prado, “Data Mining and Machine Learning to Promote Smart Cities: A Systematic Review from 2000 to 2018,” *Sustainability*, vol. 11, no. 4, 2019.
- [33] F. Correa, J. Gama, P. Correa, and L. Alves, “Data mining frequent temporal events in agrieconomic time series,” *IEEE Lat. Am. Trans.*, vol. 13, no. 7, pp. 2329–2334, 2015.
- [34] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
- [35] P. Cereda and J. Neto, “Adaptive Data Mining: Preliminary Studies,” *IEEE Lat. Am. Trans.*, vol. 12, no. 7, pp. 1258–1270, 2014.
- [36] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques,” *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, 2015.
- [37] S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, “An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives,” *Expert Syst. Appl.*, vol. 85, pp. 169–181, 2017.
- [38] J. Zhang, S. O. Williams, and H. Wang, “Intelligent computing system based on pattern recognition and data mining algorithms,” *Sustain. Comput. Informatics Syst.*, vol. 20, pp. 192–202, 2018.
- [39] Z. Ge, Z. Song, S. Ding, and B. Huang, “Data Mining and Analytics in the Process Industry: the Role of Machine Learning,” *IEEE Access*, vol. 5, pp. 20590–20616, 2017.
- [40] Y. O. Sayad, H. Mousannif, and H. Al Moatassime, “Predictive modeling of wildfires: A new dataset and machine learning approach,” *Fire Saf. J.*, vol. 104, pp. 130–146, 2019.
- [41] P. Seele, “Predictive Sustainability Control: A review assessing the potential to transfer big data driven ‘predictive policing’ to corporate sustainability management,” *J. Clean. Prod.*, vol. 153, pp. 673–686, 2017.
- [42] D. M. Rousseau, *Envisioning evidence-based management*. The Oxford Handbook of Evidence-Based Management, 2012.
- [43] S. Seuring and S. Gold, “Conducting content-analysis based literature reviews in supply chain management,” *Supply Chain Manag. An Int. J.*, vol. 17, no. 5, pp. 544–555, 2012.
- [44] W. ten Ham-Baloyi and P. Jordan, “Systematic review as a research method in post-graduate nursing education,” *Heal. SA Gesondheid*, vol. 21, pp. 120–128, 2016.
- [45] R. Pelissari, K. Manoel Infante, M. Oliveira, A. Helleno, and A. Abackerli, “A new web-based method for automatic selection of articles for systematic literature reviews,” *IEEE Lat. Am. Trans.*, vol. 15, no. 10, pp. 1926–1932, 2017.
- [46] M. V. Barros, R. Salvador, C. M. Piekarski, and A. C. de Francisco, “Mapping of main research lines concerning life cycle studies on packaging systems in Brazil and in the world,” *Int. J. Life Cycle Assess.*, pp. 1–15, 2018.
- [47] R. Pagani, J. Kovaleski, and L. Resende, “Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication,” *Scientometrics*, vol. 104, pp. 1–27, 2015.
- [48] R. T. de O. Lacerda, L. Ensslin, and S. R. Ensslin, “Uma análise bibliométrica da literatura sobre estratégia e avaliação de desempenho,” *Gestão & Produção*, vol. 19, pp. 59–78, 2012.
- [49] N. J. van Eck and L. Waltman, “Software survey: VOSviewer, a computer program for bibliometric mapping,” *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010.
- [50] M. Aria and C. Cuccurullo, “bibliometrix: An R-tool for comprehensive science mapping analysis,” *J. Informetr.*, vol. 11, no. 4, pp. 959–975, 2017.
- [51] H. Noh, Y. Jo, and S. Lee, “Keyword selection and processing strategy for applying text mining to patent analysis,” *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4348–4360, 2015.
- [52] J. Joung and K. Kim, “Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data,” *Technol. Forecast. Soc. Change*, vol. 114, pp. 281–292, 2017.
- [53] Web of Science, “Web of Science Core Collection Indexes,” 2018. [Online]. Available: https://images.webofknowledge.com/images/help/WOK/hp_database.html. [Accessed: 28-Jun-2018].
- [54] J. M. Link, “Publish or perish...but where? What is the value of impact factors?,” *Nucl. Med. Biol.*, vol. 42, no. 5, pp. 426–427, 2015.
- [55] J. Mingers and L. Leydesdorff, “A review of theory and practice in scientometrics,” *Eur. J. Oper. Res.*, vol. 246, no. 1, pp. 1–19, 2015.
- [56] R. Pagani, J. Kovaleski, and L. Resende, “Avanços na composição da Methodi Ordinatio para revisão sistemática de literatura,” *Ciência da Informação*, vol. 46, no. 2, pp. 161–187, 2017.
- [57] R. Dubey, A. Gunasekaran, S. J. Childe, S. F. Wamba, and T. Papadopoulos, “The impact of big data on world-class sustainable manufacturing,” *Int. J. Adv. Manuf. Technol.*, vol. 84, no. 1, pp. 631–645, 2016.
- [58] J. Sarkis and D. G. Dhavale, “Supplier selection for sustainable operations: A triple-bottom-line approach using a Bayesian framework,” *Int. J. Prod. Econ.*, vol. 166, pp. 177–191, 2015.
- [59] G. Egilmez, M. Kucukvar, and O. Tatari, “Sustainability assessment of U.S. manufacturing sectors: an economic input output-based frontier

- approach,” *J. Clean. Prod.*, vol. 53, pp. 91–102, 2013.
- [60] N. Moussiopoulos, C. Achillas, C. Vlachokostas, D. Spyridi, and K. Nikolaou, “Environmental, social and economic information management for the evaluation of sustainability in urban areas: A system of indicators for Thessaloniki, Greece,” *Cities*, vol. 27, no. 5, pp. 377–384, 2010.
- [61] E. H. Borgstein, R. Lamberts, and J. L. M. Hensen, “Evaluating energy performance in non-domestic buildings: A review,” *Energy Build.*, vol. 128, pp. 734–755, 2016.
- [62] J. Wang and H. Yue, “Food safety pre-warning system based on data mining for a sustainable food supply chain,” *Food Control*, vol. 73, pp. 223–229, 2017.
- [63] S. L. Ting, Y. K. Tse, G. T. S. Ho, S. H. Chung, and G. Pang, “Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry,” *Int. J. Prod. Econ.*, vol. 152, pp. 200–209, 2014.
- [64] R. Mikučionienė, V. Martinaitis, and E. Keras, “Evaluation of energy efficiency measures sustainability by decision tree method,” *Energy Build.*, vol. 76, pp. 64–71, 2014.
- [65] C. N. Madu, C. Kuei, and P. Lee, “Urban sustainability management: A deep learning perspective,” *Sustain. Cities Soc.*, vol. 30, pp. 1–17, 2017.
- [66] K. Park and G. E. O. Kremer, “Text mining-based categorization and user perspective analysis of environmental sustainability indicators for manufacturing and service systems,” *Ecol. Indic.*, vol. 72, pp. 803–820, 2017.
- [67] W. Te Liew, A. Adhitya, and R. Srinivasan, “Sustainability trends in the process industries: A text mining-based analysis,” *Comput. Ind.*, vol. 65, no. 3, pp. 393–400, 2014.
- [68] X. Wang, Z. Li, H. Meng, and J. Wu, “Identification of key energy efficiency drivers through global city benchmarking: A data driven approach,” *Appl. Energy*, vol. 190, pp. 18–28, 2017.
- [69] B. Tesfaye, S. Atique, N. Elias, L. Dibaba, S.-A. Shabbir, and M. Kebede, “Determinants and development of a web-based child mortality prediction model in resource-limited settings: A data mining approach,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 45–51, 2017.
- [70] E. Demartini, A. Gaviglio, and D. Bertoni, “Integrating agricultural sustainability into policy planning: A geo-referenced framework based on Rough Set theory,” *Environ. Sci. Policy*, vol. 54, pp. 226–239, 2015.
- [71] J. Massana, C. Pous, J. Melendez, and J. Colomer, “Identifying services for short-term load forecasting using data driven models in a Smart City platform,” *Sustain. Cities Soc.*, vol. 28, pp. 108–117, 2017.
- [72] M. Bohanec, M. Robnik-Šikonja, and M. K. Borštnar, “Decision-making framework with double-loop learning through interpretable black-box machine learning models,” *Ind. Manag. Data Syst.*, vol. 117, no. 7, pp. 1389–1406, 2017.
- [73] D. Antanasijević, V. Pocajt, M. Ristić, and A. Perić-Grujić, “A differential multi-criteria analysis for the assessment of sustainability performance of European countries: Beyond country ranking,” *J. Clean. Prod.*, vol. 165, pp. 213–220, 2017.
- [74] D. Yu and C. Fang, “The dynamics of public safety in cities: A case study of Shanghai from 2010 to 2025,” *Habitat Int.*, vol. 69, pp. 104–113, 2017.
- [75] J. Kim, J.-Y. Hyun, W. K. Chong, and S. Ariaratnam, “Understanding the effects of environmental factors on building energy efficiency designs and credits: Case studies using data mining and real-time data,” *J. Eng. Des. Technol.*, vol. 15, no. 3, pp. 270–285, 2017.
- [76] M. Panda, “Intelligent Data Analysis for Sustainable Smart Grids using Hybrid Classification by Genetic Algorithm based Discretization,” *Intell. Decis. Technol.*, vol. 11, 2017.
- [77] Y. Everingham, J. Sexton, D. Skocaj, and G. Inman-Bamber, “Accurate prediction of sugarcane yield using a random forest algorithm,” *Agron. Sustain. Dev.*, vol. 36, no. 2, p. 27, 2016.
- [78] L. Zeng, B. Wang, L. Fan, and J. Wu, “Analyzing sustainability of Chinese mining cities using an association rule mining approach,” *Resour. Policy*, vol. 49, pp. 394–404, 2016.
- [79] A. Capozzoli, G. Serale, M. S. Piscitelli, and D. Grassi, “Data mining for energy analysis of a large data set of flats,” *Proc. Inst. Civ. Eng. - Eng. Sustain.*, vol. 170, no. 1, pp. 3–18, 2017.
- [80] T.-L. Tseng and C.-C. Huang, “Sustainable service and energy provision based on agile rule induction,” *Int. J. Prod. Econ.*, vol. 181, pp. 273–288, 2016.
- [81] S. J. Rivera, B. S. Minsker, D. B. Work, and D. Roth, “A text mining framework for advancing sustainability indicators,” *Environ. Model. Softw.*, vol. 62, pp. 128–138, 2014.
- [82] M. Debeljak, A. Poljanec, and B. Ženko, “Modelling forest growing stock from inventory data: A data mining approach,” *Ecol. Indic.*, vol. 41, pp. 30–39, 2014.
- [83] S. Chi, M. Murphy, and Z. Zhang, “Sustainable Road Management in Texas: network-level flexible pavement structural condition analysis using data mining techniques,” *J. Comput. Civ. Eng.*, 2012.
- [84] V. Mithal et al., “Monitoring Global Forest Cover Using Data Mining,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 4, p. 36:1–36:24, 2011.
- [85] D. Stojanova, A. Kobler, P. Ogrinc, B. Ženko, and S. Džeroski, “Estimating the risk of fire outbreaks in the natural environment,” *Data Min. Knowl. Discov.*, vol. 24, no. 2, pp. 411–442, Mar. 2012.
- [86] M. Shaheen, M. Shahbaz, A. Guergachi, and Z. Rehman, “Mining sustainability indicators to classify hydrocarbon development,” *Knowledge-Based Syst.*, vol. 24, no. 8, pp. 1159–1168, 2011.
- [87] S. Vinodh, E. Selvan, and N. Hari Prakash, “Evaluation of sustainability using fuzzy association rules mining,” *Clean Technol. Environ. Policy*, vol. 13, 2011.
- [88] B. Ekasingh and K. Ngamsomsuke, “Searching for simplified farmers’ crop choice models for integrated watershed management in Thailand: A data mining approach,” *Environ. Model. Softw.*, vol. 24, no. 12, pp. 1373–1380, 2009.
- [89] M. Flinkman, W. Michalowski, S. Nilsson, R. Slowinski, R. Susmaga, and S. Wilk, “Use Of Rough Sets Analysis To Classify Siberian Forest Ecosystems According To Net Primary Production Of Phytomass,” *INFOR Inf. Syst. Oper. Res.*, vol. 38, no. 3, pp. 145–160, 2000.
- [90] L. Rokach and O. Maimon, *Data Mining With Decision Trees: Theory and Applications*, 2nd ed. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2014.
- [91] A. da Cruz Figueira, C. S. Pitombo, P. T. M. e Silva de Oliveira, and A. P. C. Larocca, “Identification of rules induced through decision tree algorithm for detection of traffic accidents with victims: A study case from Brazil,” *Case Stud. Transp. Policy*, vol. 5, no. 2, pp. 200–207, 2017.
- [92] N. E. I. Karabadjji, H. Seridi, F. Bousetouane, W. Dhifli, and S. Aridhi, “An evolutionary scheme for decision tree construction,” *Knowledge-Based Syst.*, vol. 119, pp. 166–177, 2017.
- [93] W. Chen, C. Xie, P. Shang, and Q. Peng, “Visual analysis of user-driven association rule mining,” *J. Vis. Lang. Comput.*, vol. 42, pp. 76–

- 85, 2017.
- [94] D. J. Prajapati, S. Garg, and N. C. Chauhan, "Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment," *Futur. Comput. Informatics J.*, vol. 2, no. 1, pp. 19–30, 2017.
- [95] M. D. Ruiz, J. Gómez-Romero, M. Molina-Solana, J. R. Campaña, and M. J. Martín-Bautista, "Meta-association rules for mining interesting associations in multiple datasets," *Appl. Soft Comput.*, vol. 49, pp. 212–223, 2016.
- [96] K. Noiva, J. E. Fernández, and J. L. Wescoat, "Cluster analysis of urban water supply and demand: Toward large-scale comparative sustainability planning," *Sustain. Cities Soc.*, vol. 27, pp. 484–496, 2016.
- [97] L. Wildgaard, "A critical cluster analysis of 44 indicators of author-level performance," *J. Informetr.*, vol. 10, no. 4, pp. 1055–1078, 2016.
- [98] R. Orsi, "Use of multiple cluster analysis methods to explore the validity of a community outcomes concept map," *Eval. Program Plann.*, vol. 60, pp. 277–283, 2017.
- [99] Y. Jin, W. Cao, M. Wu, and Y. Yuan, "Accurate fuzzy predictive models through complexity reduction based on decision of needed fuzzy rules," *Neurocomputing*, vol. 323, pp. 344–351, 2019.
- [100] H. Li, H. An, Y. Wang, J. Huang, and X. Gao, "Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network," *Phys. A Stat. Mech. its Appl.*, vol. 450, pp. 657–669, 2016.
- [101] H. Liao, M. Tang, L. Luo, C. Li, F. Chiclana, and X.-J. Zeng, "A Bibliometric Analysis and Visualization of Medical Big Data Research," *Sustainability*, vol. 10, no. 1, 2018.
- [102] D. Gu, J. Li, X. Li, and C. Liang, "Visualizing the knowledge structure and evolution of big data research in healthcare informatics," *Int. J. Med. Inform.*, vol. 98, pp. 22–32, 2017.



Jovani Taveira de Souza possui graduação em Licenciatura em Matemática (2013) pela Universidade Estadual de Ponta Grossa – UEPG, especialização em Engenharia de Produção (2014) e mestrado em Engenharia de Produção (2017), ambas pela Universidade Tecnológica Federal do Paraná – UTFPR. Atualmente é aluno de doutorado em Engenharia de Produção, também pela UTFPR. Áreas de pesquisa e interesse: Data Mining, Gestão do Conhecimento, Sustentabilidade. É pesquisador no Laboratório de Estudos em Sistemas Produtivos Sustentáveis (LESP / UTFPR-PG).



Antonio Carlos de Francisco possui mestrado em Tecnologia (1999) pela Universidade Tecnológica Federal do Paraná – UTFPR, doutorado em Engenharia de Produção (2003) pela Universidade Federal de Santa Catarina – UFSC. Atualmente é professor titular do Ensino Básico, Técnico e Tecnológico, além de professor e coordenador do Programa de Pós-Graduação em Engenharia de Produção da

UTFPR. Áreas de pesquisa e interesse: Sistemas Sustentáveis, Ciclo de Vida, Economia Circular, Qualidade de Vida e Qualidade de Vida no Trabalho, Data Mining. Coordenador do Laboratório de Estudos em Sistemas Produtivos Sustentáveis (LESP / UTFPR-PG).



Cassiano Moro Piekarski possui graduação em Engenharia de Produção em Controle e Automação (2011), mestrado em Engenharia de Produção (2013) e doutorado em Engenharia de Produção (2015), ambas pela Universidade Tecnológica Federal do Paraná – UTFPR. Atualmente é professor do Programa de Pós-Graduação em Engenharia de Produção (PPGEP/UTFPR-PG) e de magistério superior do Departamento de Engenharia de Produção da UTFPR, Campus Ponta Grossa. Áreas de pesquisa e interesse: Avaliação do Ciclo de Vida, Gerenciamento do Ciclo de Vida, Economia Circular, Bioenergia, Sistemas Sustentáveis. É pesquisador no Laboratório de Estudos em Sistemas Produtivos Sustentáveis (LESP / UTFPR-PG).



Guilherme Francisco do Prado possui graduação em Administração pelo Centro de Ensino Superior dos Campos Gerais – CESCAGE (2015), especialização em Engenharia de Produção (2016) pela Universidade Tecnológica Federal do Paraná – UTFPR. Atualmente é aluno de mestrado em Engenharia de Produção pela UTFPR. Áreas de pesquisa e interesse: Sustentabilidade e Desempenho econômico. É pesquisador no Laboratório de Estudos em Sistemas Produtivos Sustentáveis (LESP / UTFPR-PG).



Leandro Gasparello de Oliveira possui graduação em Administração (2009) pela Faculdade União, especialização em Engenharia de Produção (2015) e mestrado em Engenharia de Produção (2018), ambas pela Universidade Tecnológica Federal do Paraná – UTFPR. Atualmente é professor na Faculdade Unopar. Áreas de pesquisa e interesse: Energias Alternativas, Geração de Biogás e Biometano, Avaliação do Ciclo de Vida e Sustentabilidade. É pesquisador no Laboratório de Estudos em Sistemas Produtivos Sustentáveis (LESP / UTFPR-PG).